**Chapter 1**

**INTRODUCTION**

**Motivation:**

We have chosen to work with twitter because there are hundreds of thousands of users around the globe who have figured out the value of Twitter and are using it in a variety of ways. If you’ve been wondering, “What is Twitter used for? ”

**Twitter Is Used For Connecting People:**

First, Twitter is used to connect people with the same interests. As the twitter homepage suggests, the social platform can be used to, “Connect with your friends — and other fascinating people. Get in-the-moment updates on the things that interest you.”This process of connecting people who are complete strangers can be done with the use of hashtags. Hashtags, which are denoted with the “#” prefix, are added to Tweets so members of the community can share in the conversation. Users can even use a website like hashtag.org to find topics that interest them. They can then use those hashtags to join in the conversations taking place on the subject, ultimately helping to build online communities based on content.

**Twitter Is Used to Share Information In Real-Time**

When major events take place, Twitter lights up with Tweets. We’ve seen this happen in a variety of ways, including when popular television shows or award shows are on, or when significant events are unfolding. For instance, when Barack Obama was re-elected as President of the United States in 2012, the event received 327,000 Tweets per minute.

According to The Next Web, the 2014 Brazil-China World Cup game became the most Tweeted sports event in history, which included 16.4 million Tweets being sent during the game.

Due to the nature of Twitter, and the rampant accessibility of the social platform through smartphones and tablets, users can Tweet about their experiences as soon as they happen — making Twitter a very powerful social tool.

**Twitter Is Used for Marketing In Business**

There are different ways Twitter can be used by businesses.First, let’s consider web-only businesses that generate revenue solely through advertisements. These properties can Tweet about the content they provide or the activities they’re involved in to drive more traffic to their website, ultimately generating more revenue for them. To build subscribers, the company could use hashtags related to its content to find its audience members.

Other companies including business-to-business or business-to-consumer can spread its content or product information through Twitter in the same way.

Content-based business like publishers who have a lot of written content on their websites uses Twitter for search engine optimization (SEO) purposes. Although Matt Cutts of Google’s Web Team has specifically said that social signals from Twitter and Facebook don’t play part in Google’s ranking algorithm, Tweeting about articles and web pages helps drive more traffic to them, ultimately creating the possibility of better rank.



Fig 1.1: Twitter is used for business purpose

In addition to the organic usage of Twitter, businesses on Twitter can pay for Twitter advertisements. Companies that do advertise on Twitter have the option of targeting audiences through keywords, demographics, location, and interests. Accounts and Tweets can also be promoted, which brings them in front of users who wouldn’t necessarily see the content in any other way. Users that opt for promoted Tweets don’t have to pay unless the content is retweeted, replied to, favorited or clicked. Promoted Account users don’t have to pay unless people follow  the account.

Twitter is also used by businesses for branding purposes, bringing information of a brand out to the masses easily.

**Twitter Is Used as an Educational Tool**

In a world that’s always changing, new forms of education are constantly developing. With an extremely digital environment spanning the globe, educators are teaching the relevance of Twitter to their students.

November Learning cites three specific usages of Twitter in the educational realm:

- Using Twitter to facilitate authentic conversation with students.

- Using Twitter to connect students with real-world problems.

- Using Twitter to expand the boundaries of learning that traditional textbooks cannot do.

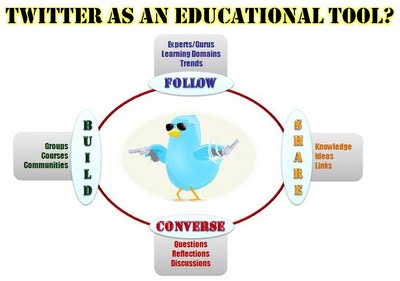


Fig 1.2: Twitter as an Education Tool

**Domain Introduction:**

This project of analyzing hastags of tweets comes under the domain of “**Pattern Classification**” and “**Data Mining**”. Both of these terms are very closely related and intertwined, and they can be formally defined as the process of discovering “useful” patterns in large set of data, either automatically (unsupervised) or semi-automatically (supervised). The project would heavily rely on techniques of “Natural Language Processing” in extracting significant patterns and features from the large data set of tweets and on “Machine Learning” techniques for accurately classifying individual unlabelled data samples (tweets) according to whichever pattern model best describes them.

Twitter based features are more informal and relate with how people express themselves on online social platforms and compress their sentiments in the limited space of 140 characters offered by twitter. They include twitter hashtags, retweets, word capitalization, word lengthening question marks, presence of url in tweets, exclamation marks, internet emoticons and internet shorthand/slangs.

**Data mining** is the process of discovering patterns in large [data sets](https://en.wikipedia.org/wiki/Data_set) involving methods at the intersection of [machine learning](https://en.wikipedia.org/wiki/Machine_learning), [statistics](https://en.wikipedia.org/wiki/Statistics), and [database systems](https://en.wikipedia.org/wiki/Database_system). Data mining is an [interdisciplinary](https://en.wikipedia.org/wiki/Interdisciplinary) subfield of [computer science](https://en.wikipedia.org/wiki/Computer_science) with an overall goal to extract information (with intelligent method) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD .Aside from the raw analysis step, it also involves database and [data management](https://en.wikipedia.org/wiki/Data_management) aspects, [data pre-processing](https://en.wikipedia.org/wiki/Data_pre-processing), [model](https://en.wikipedia.org/wiki/Statistical_model) and [inference](https://en.wikipedia.org/wiki/Statistical_inference) considerations,interestingness metrics.

The term "data mining" is in fact a [misnomer](https://en.wikipedia.org/wiki/Misnomer), because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (*mining*) of data itself.

Rest assured, you need not track every single hash tag on Twitter. However, it is essential you narrow it down to the most trending hashtags in your industry, hashtags used by influences and branded hashtags. To simplify this task, hashtags can be divided into two broad categorie

**Short Term:**

As the name indicates, these hashtags are popular for a short period and are typically based on current events. Most hashtags on Twitter are in fact temporary, which means they are only used for events currently taking place or for hot stories. Your brand needs to track these hashtags for social media contests, conferences and breaking industry news. This is essential if you want your brand to be ahead of competitors.

**Long Term:**

These hashtags are commonly used and generic in the world of social media. For example, hashtags, such as #NoFilter or #TBT (throwback Thursday), are consistently used on Instagram. However, not all these hashtags may be relevant to your industry.

**Chapter 2**

**LITERATURE REVIEW**

**2.1 Twitter Research:**

Twitter was launched in 2006 as a microblogging platform that facilitated users in sharing and consuming information about day-to-day happenings and opinions on topics. It was a unique product during the time of its introduction due to its character limit on user posts (users are allowed to post tweet messages within 140 characters limit). Twitter has become immensely popular (Rank 10 in Alexa web rankings3 , Rank 3 among social networking sites4 ) which has to lead to regional spinoffs such as Sina Weibo5 . Noticing the dynamics around the interactions in twitter, academic research in twitter started in 2008- 2009(Krishnamurthy et al., 2008). Twitter research has been surveyed and summarized in (Cheong & Lee, 2010; Cheong & Ray, 2011; Williams, Terras, & Warwick, 2013). Research has furthered in different directions with varied focus such as organising information (Sriram, Fuhry, Demir, Ferhatosmanoglu, & Demirbas, 2010), understanding trends and convergence events from a communications perspective(Lin et al., 2012), usage of twitter data in practical applications (ex: governments, activism) (Bruns & Burgess, 2011), cross-application of twitter data (in cross-platform recommendations(Abel, Herder, Houben, & Henze, 2011)) and traditional computer science oriented focus on information retrieval (Magnani, Montesi, Nunziante, & Rossi, 2011) and semantics(Abel, Celik, Houben, & Siehndel, 2011)

**References:**

Cheong, M., & Lee, V. (2010). Dissecting Twitter: A Review on Current Microblogging Research and Lessongs from Related Fields. From Sociology to Computing in Social Networks (pp. 343–362). Springer Vienna.

**2.2 Tweet Classification:**

The two high level entities in twitter are User and Message (Cheong & Lee, 2010). Recent research has introduced two additional entities Technology and Concept (the central topic being addressed in the tweet)(Williams et al., 2013). This classification scheme has been used in studying twitter data. On the topic of tweet6 classification, past research has identified many categories which differ based on the method of classification, amount of data, period of data and frame of reference. The categories identified by earlier research studies are presented in Table 1 (a&b). Categories such as News, Information sharing, Events, Opinions and Promotions appear to be common across the schemes. The variation in schemes is mainly due to the vocabulary used for naming the categories and the purpose of classification. There has been a lack of consolidation across studies except for the work of Dann(2010) where four earlier classification schemes have been combined to form a new scheme with six generic categories. It is to be noted that all these classification attempts have not used hashtag as the frame of reference.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Java et al(2007) | Jansen et al (2009) | Honeycutt & Herring (2009) | Pear Analytics (2009) | Horn (2010) |
| Conversations URL sharing News reporting Daily chatter | Info seeking Info providing Comment/Sentiment | About addressee Advertise Exhort Info for others Info for self Meta-commentary Media use Express opinion Other's experience Self experience Solicit info Other miscellaneous | Mainstream News Spam Self-promotion of businesses Babble Conversations Pass-along messages (retweets) | C1: News, Events, Company C2: Factual, Opionated |

Table 2.1:Classification Schemes from Previous Twitter Studies(2007-2010)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sriram et al (2010) | Dann (2010) | Sandra et al (2010) | Naaman et al (2010) | Rosa (2011) |
| News Opinions Deals Table 2.1:Classification Schemes from Previous Twitter Studies(2007-2010)Events Private Messages | Conversational Pass along News Status Phatic Spam | Movies Books Music Apps Games | Info sharing Self-promotionOpinions/Complaints Statements & Random Thoughts Me now Question to followers Presence Maintenance Anecdote | News Sports Science & Technology Entertainment Money/Business Just for Fun |

Table 2.2:Classification Schemes from Previous Twitter Studies (2010-2011)

**References:**

Dann, S. (2010). Twitter content classification. First Monday, 15(12). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2745/2681

**2.3 Hashtag Studies**

Hashtag is a keyword which starts with the symbol ‘#’. It is mainly used for categorizing content and joining conversations on various topics (Huang, Thornton, & Efthimiadis, 2010). Hashtags serve the same purpose as tags made famous by web 2.0 services such as Flickr7 and Delicious8 . Even though, users need not necessarily add hashtags to their tweets, it is generally observed that regular users add hashtags to most of their tweets. Global political events are represented in twitter through hashtags, some of the popular ones include #occupy, #OWS and#Syria. #sghaze was quite popular in Singapore during the period of August 2013.The analysis of behavior around hashtags was done as part of an earlier study on tagging (Huang et al., 2010). The authors use statistical measures such as standard deviation, skewness and kurtosis to study the popularity of hashtags and the scenarios in which hashtags gain traction. Pöschko (2011) performed an exploratory study on hashtags by analysing 29 million tweets which involved tweet classification, studying hashtag co-occurrences, part-of-speech tagging and SNA based clustering thereby highlighting different ways of dissecting the tweet data to gain insights. Yang et al (2012) developed a machine learning model to predict the future adoption of hashtags by users, by combining the two use-cases under which a hashtag is used by users.

**2.4 Data Collection:**

To track and monitor different datasets, most studies began with collecting the desired datasets from twitter, and applied filtering techniques to remove redundant data or spam tweets. Then parsed the data into a structured form. There are 4 primary ways of acquiring Twitter data (and I’m not including “cutting and pasting” from the Twitter website!):

Retrieve from the Twitter public API.

* Find an existing Twitter dataset.
* Purchase from Twitter.
* Access or purchase from a Twitter service provider.

Let’s explore each of these.

**1. Retrieve from the Twitter public API**

API is short for “Application Programming Interface” and in this case is a way for software to access the Twitter platform (as opposed to the Twitter website, which is how humans access Twitter). While supporting a large number of functions for interacting with Twitter, the API functions most relevant for acquiring a Twitter dataset include:

* Retrieving tweets from a user timeline (i.e., the list of tweets posted by an account)
* Searching tweets
* Filtering real-time tweets (i.e., the tweets as they are passing through the Twitter platform upon posting)

While you can write your own software for accessing the Twitter API, a number of tools already exist. They are quite varied in their capabilities and require different levels of technical skills and infrastructure. These include:

* Software libraries (e.g., Tweepy for Python and rtweet for R)
* Command line tools (e.g., Twarc)
* Web applications (e.g., DMI-TCAT and our very own Social Feed Manager)
* Plugins for popular analytic packages (e.g., NVIVO, NodeXL for Excel, and TAGS for Google Sheets)

Some of these tools are focused on retrieving tweets from the API, while others will also do analysis of the Twitter data. For a more complete list, see the Social Media Research Toolkit from the Social Media Lab at Ted Rogers School of Management, Ryerson University.

Note when selecting a tool that some may only support part of the Twitter API for retrieving tweets, most commonly, search. Further, some tools may be designed to support one-time retrieval from the Twitter API, while others support retrieval on an ongoing basis. (For example, Social Feed Manager allows you to specify a schedule for recurring data collection.)

What all of these tools share in common is that they use Twitter’s public API. The Twitter public API has a number of limitations that you should be aware of:

* Access to historical tweets is extremely limited. You can retrieve the last 3,200 tweets from a user timeline and search the last 7-9 days of tweets.
* Access to current tweets is limited. Depending on how broad your filter is, the API may not return all tweets.
* Twitter may sample or otherwise not provide a complete set of tweets in searches.

### 2. Find an existing Twitter dataset

One way to overcome the limitations of Twitter’s public API for retrieving historical tweets is to find a dataset that has already been collected and satisfies your research requirements. For example, here at GW Libraries we have proactively built collections on a number of topics including Congress, the federal government, and news organizations.

Twitter’s Developer Policy (which you agree to when you get keys for the Twitter API) places limits on the sharing of datasets. If you are sharing datasets of tweets, you can only publicly share the ids of the tweets, not the tweets themselves. Another party that wants to use the dataset has to retrieve the complete tweet from the Twitter API based on the tweet id (“hydrating”). Any tweets which have been deleted or become protected will not be available.

DocNow’s Hydrator is a useful tool for retrieving tweets from the Twitter API based on tweet id. Note that Twitter places rate limits on hydrating (as it does on most API functions) so this may take some amount of time depending on the size of the dataset.

A number of individuals and organizations have publicly posted Twitter datasets, e.g., in a dataset repository or on a website. For example, we posted our 280 million tweet dataset from the 2016 U.S. presidential election on Harvard’s Dataverse. Deen Freelon has published the 40 million tweet dataset for the “Beyond the Hashtags: #Ferguson, #Blacklivesmatter, and the Online Struggle for Offline Justice” report on his website. The DocNow Catalog provides a listing of publicly available Twitter datasets.

Twitter’s Developer Policy is generally interpreted as allowing sharing of tweets locally, i.e. within an academic institution. For example, we share the datasets we have collected at GW Libraries with members of the GW research community (but when sharing outside the GW community, we only share the tweet ids). However, only a small number of institutions proactively collect Twitter data – your library is a good place to inquire.

Another option for acquiring an existing Twitter dataset is TweetSets, a web application that I’ve developed. TweetSets allows you to create your own dataset by querying and limiting an existing dataset. For example, you can create a dataset that only contains original tweets with the term “trump” from the Women’s March dataset.

### 3. Purchase from Twitter

You can purchase historical Twitter data directly from Twitter, using the Historical PowerTrack enterprise product.

Historical Twitter data was previously available from Gnip, a data service provider purchased by Twitter. Gnip has now been folded into Twitter. The way this used to work is that you provided a set of query terms and other limiters and a Gnip sales rep replied with a cost estimate. With recent changes, the process is less clear.

For filtering tweets, the Historical Powertrack offers a number of enhancements over the public Twitter API. This includes additional filter operators and tweet enhancements (e.g., profile location and unshortened URLs).

When considering purchasing tweets, you should be aware that it is not likely to be a trivial amount of money. The cost depends on both the length of the time period and the number of tweets; often, the cost is driven by the length of the time period, so shorter periods are more affordable. The cost may be feasible for some research projects, especially if the cost can be written into a grant. Further, I am not familiar with the conditions placed on the uses / sharing of the purchased dataset. Nonetheless, this is likely to be as complete a dataset as it is possible to get.

### 4. Access or purchase from a Twitter service provider

A number of commercial and academic organizations act as Twitter service providers, usually for a fee. These services provide:

* Access to Twitter data.
* Value-added services for the Twitter data, such as coding, classification, analysis, or data enhancement. If you are not using your own tools for analysis, these value-added services may be extremely useful for your research (or they may be used in combination with your own tools).

Twitter data options available from a service provider generally include one or more of the following types (available at different costs):

* Data from the public Twitter APIs. This obviously comes with the limitations described previously with the public Twitter APIs, but will be less costly than the other Twitter data options.
* Data from the enterprise Twitter APIs, which have access to all historical tweets. Like purchasing data directly from Twitter, the cost will depend on factors such as the number of tweets and the length of the time period. DiscoverText offers this type of data acquisition.
* Datasets built by querying against an existing set of historical tweets. The service provider will have an arrangement with Twitter that will provide them with access to the “firehose” of all tweets to build this collection. Crimson Hexagon offers this type of data acquisition.

Twitter service providers generally provide reliable access to the APIs, with redundancy and backfill. This means that you will not miss tweets because of network problems or other issues that might occur when using a tool to access the APIs yourself. Note, also, that some service providers can provide data from other social media platforms, such as Facebook.

Despite what the sales representative may tell you, most Twitter service providers’ offerings focus on marketing and business intelligence, not academic research. The notable exception is DiscoverText, which is focused primarily on supporting academic researchers. DiscoverText allows you to acquire data from the public Twitter APIs, as well as purchasing historical tweets. Within the DiscoverText platform, tweets can be filtered, coded, and classified (using machine learning).Crimson Hexagon focuses on marketing, but also supports academic research. Social Data Analytics at Boston University is a new entry in the academic field.

There are some limitations of Twitter service providers that you should be aware of. Whether these limitations are significant will depend on your research requirements.

First, when considering a Twitter service provider, it is important to know whether you are able to export your dataset from the service provider’s platform. (All should allow you to export reports or analysis.) For most platforms, export is limited to 50,000 tweets per day.If you need the raw data to perform your own analysis or for data sharing, this may be an important consideration. Second, while the value-added services offered by a Twitter service provider may be very powerful and not require technical skill to use, they are generally a “black box”. So, for example, if a service provider performs bot detection, you may not know which bot detection algorithm is being used.

**Chapter 3**

**ANALYSIS**

**3.1 Existing System:**

Message Classification for Twitter Data is the existing system.

**Description of the Data:**

The Data that are used were collected manually from six different fields which are Sports, Business, Technology, Entertainment, Politics and Education. Table 1 shows the distribution of messages.

|  |  |
| --- | --- |
| **Topics (Class)** | **Number of Tweets** |
| Sports | 303 |
| Entertainment | 204 |
| Business | 230 |
| Politics | 206 |
| Education | 152 |
| Technology | 167 |

Table 3.1: Diversity of tweet data set

**Pre-processing of Data:**

The following pre-processing steps were done o Removal of Punctuation and Stop words o Remove the Word containing ‘#’, ‘@’ and ‘http’ or ‘https’ which means the trending topics, user name and URL are removed from tweets o Numerical character and special character are removed from the tweets Each word of the tweets are lemmatized .

**Data Representation**

We have used bag-of-word technique for data representation. The bagof-words model is a simplifying representation used in language processing. In this model, a text is represented as the bag of its words, disregarding grammar and even word order but keeping multiplicity

.

**Disadvantages:**

* Humans are prone to error, while computers are not. The more work you dedicate to machines, the more accurate your records will be.
* Machines are faster than humans. This gives you the ability to perform many more business functions, from inventory review to creating labor charts, in a shorter amount of time.
* One person can perform many functions at one time using automatic data entry instead of the manual method. This means money saved in labor costs across the board when you use machines for basic data entry and calculation.

**3.2 Proposed System:**

In this project a web page is displayed where we have to enter a hashtag. Then it analysis the hashtag we have entered and redirects to a web page which contains fulltext ,retweet count, favourite count, user id of that particular hash tag and it is sorted based on most favouite, most users, full text, most retweets.

### HOW HASH TAG ANALYSIS IS BENEFICIAL FOR ALL

### SOCIAL MEDIA MARKETERS

Social Media Marketers can benefit a lot by tracking hashtags for their social campaigns, hashtags related to their domain. Below are some of key benefits of hashtag analytics for marketers.

**1.FIND INFLUENCERS FOR LONG TERM MANAGEMENT**:

Social media marketers also benefit from hashtag analytics because it enables them to identify the influencers in the niche. By using the domain hashtag like #socialselling, social media marketers can hunt down other influencers and engage in conversation with them. This can potentially help people increase their number of followers on Twitter and who doesn’t want that?

**BRAND MANAGERS**

Brand Managers run customized hashtag campaigns. They need to track and analyze campaigns of their brand as well competitor brands. Below are key benefits they can achieve for their brand success.

**1.USE HASH TAG ANALYTICS TO MEASURE REACH**

Brand managers can use hashtag analytics to manage their hashtag campaigns, measure reach and impressions. This gives them an idea of how well they are doing and drive their campaign to success.

**2.ANAYLSE THE POPULAR HASHTAGS OF THEIR DOMAIN FOR BRANDING**

Brand managers can use hashtag analytics to explore and analyze popular hashtags. You can share these Tweets to increase brand engagement. Not to mention, popular hashtags can be incorporated while sharing posts to increase brand awareness. In this example, the brand is using popular hashtags to advertise their brand.

**EVENT HOSTS**

Event hosts generally use customized hashtag for their event. They can analyze their hashtag to find insights of their hashtag and find useful metrics like below.

**1.FIND POPULAR TWEETS OF YOUR EVENT**

Event hosts can use hashtag analytics to find popular tweets. Quite recently, the #SMMW17 was a success with guests and event host frequently tweeting about the event.

**2.FIND ENGAGING TOPICS AUDIENCE ARE DISCUSSING**

Event hosts can use hashtag analytics to find engaging topics that are trending on social media. This will give a list of new ideas for next events to increase your audience as well engagement.

**3.FIND ENGAGING SPEAKERS OF EVENT**

Finding engaging speakers for your show can increase rating. Fortunately, this task has become easy by using hashtag analytics. Event hosts can easily find speakers according to their niche. Bonus points if the speaker has a Twitter account with thousands of followers that can serve as your target audience.

**TV CHANNELS**

TV channels also host live debates with customized hashtags that can be tracked to measure success of their programs. Below are some key metrics Channels can benefit using hashtag analytics.

**1.IMPROVE CONTENT OF THEIR DEBATE**

TV channels can use hashtag analytics to track online discussions so that can generate trending topics for their TV Channel debate. This is an excellent way to improve content and show customers what they really want to watch.Hashtag analytics is a great way to conduct research and improve content. This will ensure you always have new and trending ideas to captivate the attention of your viewers..

**2.SEE HOW WELL YOUR SHOWS ARE DOING**

Analyzing hashtags via hashtag analytics is a fool-proof way of seeing how well your shows are doing. The formula is pretty simple, [[1]](#footnote-2)the more people are talking about it, the more viewers you have.

**SPORTS HASHTAGS**

Popular sports are also come up with their own hashtag which audience use for their opinion. By analyzing these hashtags they can learn a lot from them.

**1.FIND OUT WHO THE AUDIENCE IS SUPPORTING**

Hashtag analytic tools can be used to find out who the audience or general public is supporting. Too often, brands also pick up on trending hashtags for sporting events to increase engagement.

**2.FIND THE INFLUENCER TO PROMOTE YOUR SPORT TEAM**

You can effectively use hashtag analytics to see find influencers to support your team. Finding someone with a huge Twitter following can work in your favor since it will result in more support for your team, increasing their value.

**3.3 FEASIBILITY STUDY**

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out.This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are:

Three key considerations involved in the feasibility analysis are:

* Economical Feasibility
* Technical Feasibility
* Social Feasibility

**Economical Feasibility**

It is carried out to check the economic impact that the system will have on the organization. Hashtags can help accounts gain followers, improve their reputations and create branded marketing campaigns. Today, hashtags can even “make or break” a social media post, with some sources reporting that tweets without hashtags receive two times less engagement than tweets with hashtags. That’s why it’s important for economic development organizations (EDOs) to research and utilize the best hashtags possible in their digital economic development marketing efforts.

To find the most relevant Twitter hashtags, here are four simple methods to consider.

* Go Straight to the Source.
* Use third party websites.
* Adjust your twitter trend settings.

**Technical Feasibility**

It is carried out to check the technical feasibility, that is, the technical requirements of the system. The platforms we have used in this project i.e python are platform independent. Bottle server page, HTML, CSS, twitter API , browser are used to develop this project. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing feasibility.

**Social Feasibility**

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. Twitter hashtag analysis is very helpful for the users. Rather than searching about the particular hashtag manually it is very simple when the users use this hashtag analysis. Using this hashtag analysis they can find the hash tag details with in the seconds. Most of their time will be saved and the business purpose can know which company has the more followers etc.

**Chapter 4**

**SOFTWARE REQUIREMENT SPECIFICATIONS**

**4.1 FUNCTIONAL REQUIREMENTS**

The system after careful analysis has been identified to be presented with the following requirements.

1. Design storage, control server and VM nodes.
2. Collection of software analysis features (process memory usage, peak memory usage, number of threads and number of handles.
3. Collection of network analysis features (packets address pair)
4. Generate statistical Meta features.
5. Configure Meta features on virtual machines.
6. Save software analysis features into system analysis engine.
7. Save network analysis features into network analysis engine.
8. Detection of different features based on meta features.
9. Classification of malware based on behavior.
10. Optimize the detector.

**4.2 NON-FUNCTIONAL REQUIREMENTS**

## 1. Usability

## Usability is the ease of use and learns ability of a human-made object. The object of use can be a software application, website, book, tool, machine, process, or anything a human interacts with. A usability study may be conducted as a primary job function by a usability analyst or as a secondary job function by designers, technical writers, marketing personnel, and others.

## 2. Reliability

## The probability that a component part, equipment, or system will satisfactorily perform its intended function under given circumstances, such as environmental conditions, limitations as to operating time, and frequently and thoroughness of maintenance for a specified period of time.

**3. Performance**

Accomplishment of a given task measured against preset standards of accuracy, completeness, cost, and speed.

## 4. Supportability

To which the design characteristics of a stand by or support system meet the operational requirements of an organization.

## 5. Implementation

Implementation is the realization of an application, or execution of a plan, idea, model, design, specification, standard, algorithm, or policy.

**4.3 SYSTEM REQUIREMENTS**

**SOFTWARE REQUIREMENTS**

* Operating system : Windows 7.
* Coding Language : Python.
* Tool : Python 3.6
* Front end : Html ,CSS.
* Back end : Python.
* Server : web based server.
* API : Twitter API.
* Command line tool : twarc.

**HARDWARE REQUIREMENTS**

* System : Pentium Dual Core.
* Hard Disk : 45 GB.
* Monitor : 15’’ LED .
* Input Devices : Keyboard, Mouse.
* RAM : 512MB

**Python technology:**

Python is an [interpreted](https://en.wikipedia.org/wiki/Interpreted_language) [high-level programming language](https://en.wikipedia.org/wiki/High-level_programming_language) for [general-purpose programming](https://en.wikipedia.org/wiki/General-purpose_programming_language). Created by [Guido van Rossum](https://en.wikipedia.org/wiki/Guido_van_Rossum) and first released in 1991.

An interpreter is a kind of program that executes other programs. When you write Pythonprograms , it converts source code written by the developer into intermediate language which is again translated into the native language / machine language that is executed.Compilation is simply a translation step, and byte code is a lower-level, and platform-independent , representation of your source code. Roughly, each of your source statements is translated into a group of byte code instructions. This byte code translation is performed to speed execution byte code can be run much quicker than the original source code statements.The Virtual Machine is the runtime engine of Python and it is always present as part of the Python system, and is the component that truly runs the Python scripts . Technically, it's just the last step of what is called the Python interpreter.

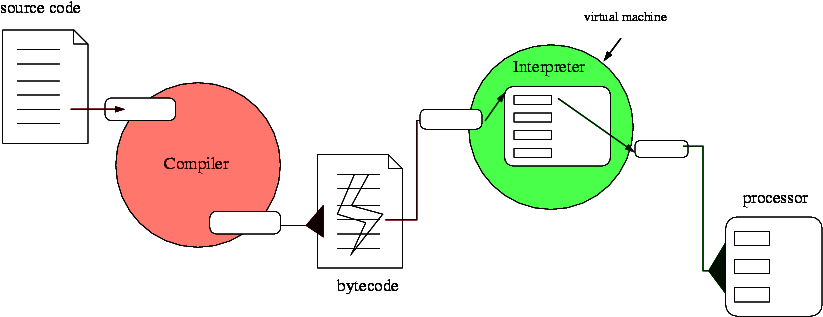


Fig 4.1: Compiling and interpreting source code

## Features of Python

**1.Simple:**

Python is a simple and minimalistic language. Reading a good Python program feels almost like reading English (but very strict English!). This pseudo-code nature of Python is one of its greatest strengths.

**2.Easy to Learn:**

As you will see, Python is extremely easy to get started with. Python has an extraordinarily simple syntax as already mentioned.

**3.Free and Open Source:**

Python is an example of a FLOSS (Free/Libre and Open Source Software). In simple terms, you can freely distribute copies of this software, read the software's source code, make changes to it, use pieces of it in new free programs, and that you know you can do these things. FLOSS is based on the concept of a community which shares knowledge. This is one of the reasons why Python is so good - it has been created and improved by a community who just want to see a better Python.

**4.High-level Language:**

When you write programs in Python, you never need to bother about low-level details such as managing the memory used by your program.

**5.Portable:**

Due to its open-source nature, Python has been ported (i.e. changed to make it work on) to many platforms. All your Python programs will work on any of these platforms without requiring any changes at all. However, you must be careful enough to avoid any system-dependent features.

**6.Interpreted**

When you use an interpreted language like Python, there is no separate compilation and execution steps. You just run the program from the source code. Internally, Python converts the source code into an intermediate form called bytecodes and then translates this into the native language of your specific computer and then runs it. All this makes using Python so much easier. You just run your programs - you never have to worry about linking and loading with libraries, etc. They are also more portable this way because you can just copy your Python program into another system of any kind and it just works!

**7.Object Oriented:**

Python supports procedure-oriented programming as well as object-oriented programming. In procedure-oriented languages, the program is built around procedures or functions which are nothing but reusable pieces of programs. In object-oriented languages, the program is built around objects which combine data and functionality. Python has a very powerful but simple way of doing object-oriented programming, especially, when compared to languages like C++ or Java.

**8.Extensible:**

If you need a critical piece of code to run very fast, you can achieve this by writing that piece of code in C, and then combine that with your Python program.

**9.Embeddable:**

You can embed Python within your C/C++ program to give scripting capabilities for your program's users.

**10.Extensive Libraries:**

The Python Standard Library is huge indeed. It can help you do various things involving regular expressions, documentation generation, unit testing, threading, databases, web browsers, CGI, ftp, email, XML, XML-RPC, HTML, WAV files, cryptography, GUI(graphical user interfaces) using Tk, and also other system-dependent stuff. Remember, all this is always available wherever Python is installed. This is called the "batteries included" philosophy of Python.

Besides the standard library, there are various other high-quality libraries such as the Python Imagining library which is an amazingly simple image manipulation library.

**Hyper Text Markup Language:**

Hypertext Markup Language (HTML), the languages of the World Wide Web (WWW), allows users to produces Web pages that include text, graphics and pointer to other Web pages (Hyperlinks).

HTML is not a programming language but it is an application of ISO Standard 8879, SGML (Standard Generalized Markup Language), but specialized to hypertext and adapted to the Web. The idea behind Hypertext is that instead of reading text in rigid linear structure, we can easily jump from one point to another point. We can navigate through the information based on our interest and preference. A markup language is simply a series of elements, each delimited with special characters that define how text or other items enclosed within the elements should be displayed. Hyperlinks are underlined or emphasized works that load to other documents or some portions of the same document.

**CSS:**

Cascading Style Sheets (CSS) is a [style sheet language](https://en.wikipedia.org/wiki/Style_sheet_language) used for describing the [presentation](https://en.wikipedia.org/wiki/Presentation_semantics) of a document written in a [markup language](https://en.wikipedia.org/wiki/Markup_language) like [HTML](https://en.wikipedia.org/wiki/HTML). CSS is designed to enable the separation of presentation and content, including [layout](https://en.wikipedia.org/wiki/Page_layout), [colors](https://en.wikipedia.org/wiki/Color), and [fonts](https://en.wikipedia.org/wiki/Typeface). This separation can improve content [accessibility](https://en.wikipedia.org/wiki/Accessibility), provide more flexibility and control in the specification of presentation characteristics, enable multiple [web pages](https://en.wikipedia.org/wiki/Web_page) to share formatting by specifying the relevant CSS in a separate .css file, and reduce complexity and repetition in the structural content.

**Twitter API:**

Having access to the Twitter API can help you manage your social media accounts, and allow you to mine social media for data. This can be useful for brand promotion if you represent a business or an organization, and it can be enjoyable and entertaining for individual users and hobbyist programmers.

## Step 1 :Create Your Twitter Application

Let’s go through the process of creating a Twitter application and retrieving your API access keys and tokens. These tokens are what will allow you to authenticate any applications you develop that work with Twitter. As mentioned in the prerequisites, you’ll need a valid phone number in order to create applications using Twitter.

Open up your browser and visit https://apps.twitter.com/ then log in using your Twitter account credentials. Once logged in, click the button labeled **Create New App**.

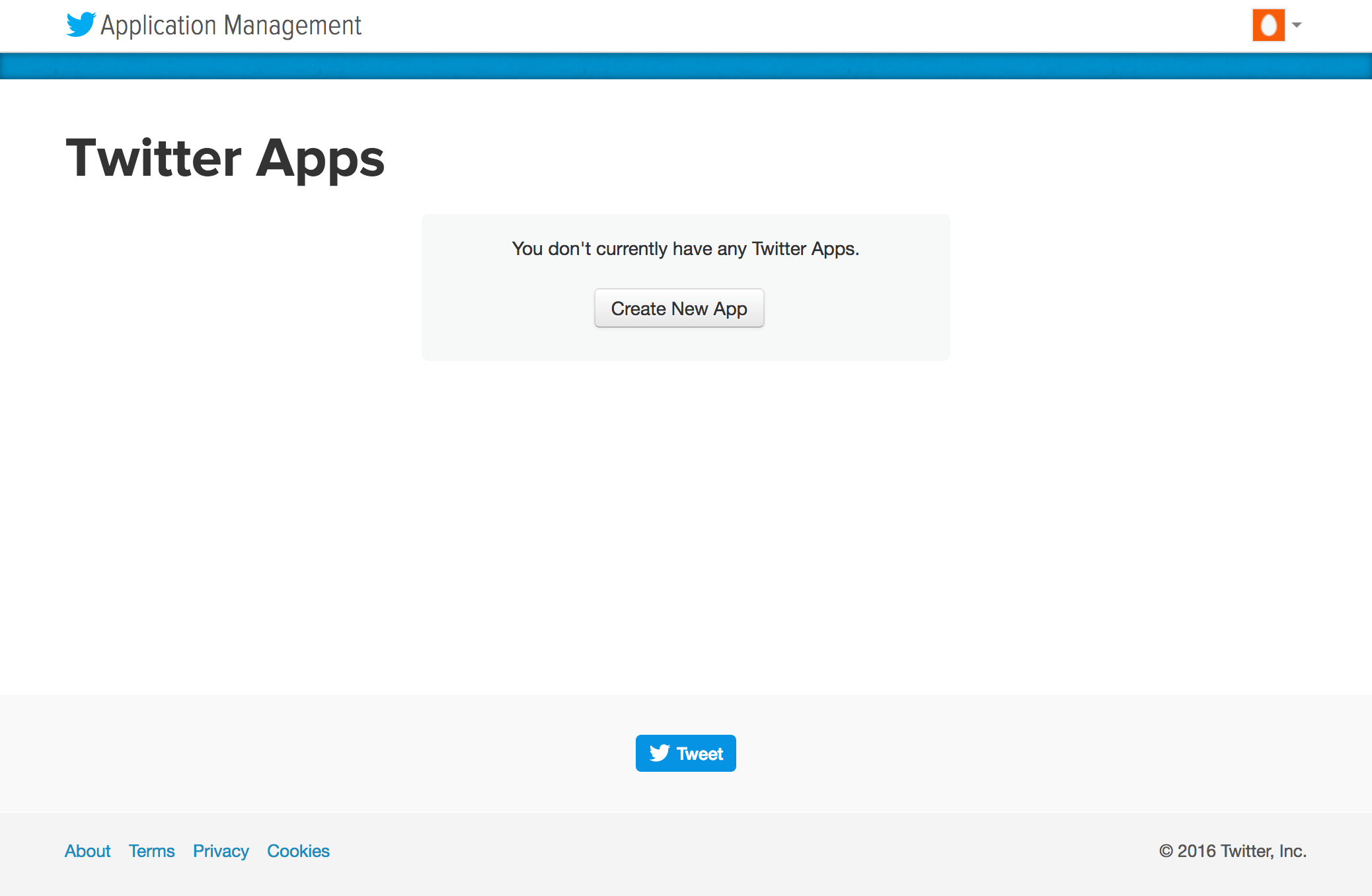
****

Fig 4.2 Creating twitter application

**Description:**

You will now be redirected to the application creation page.

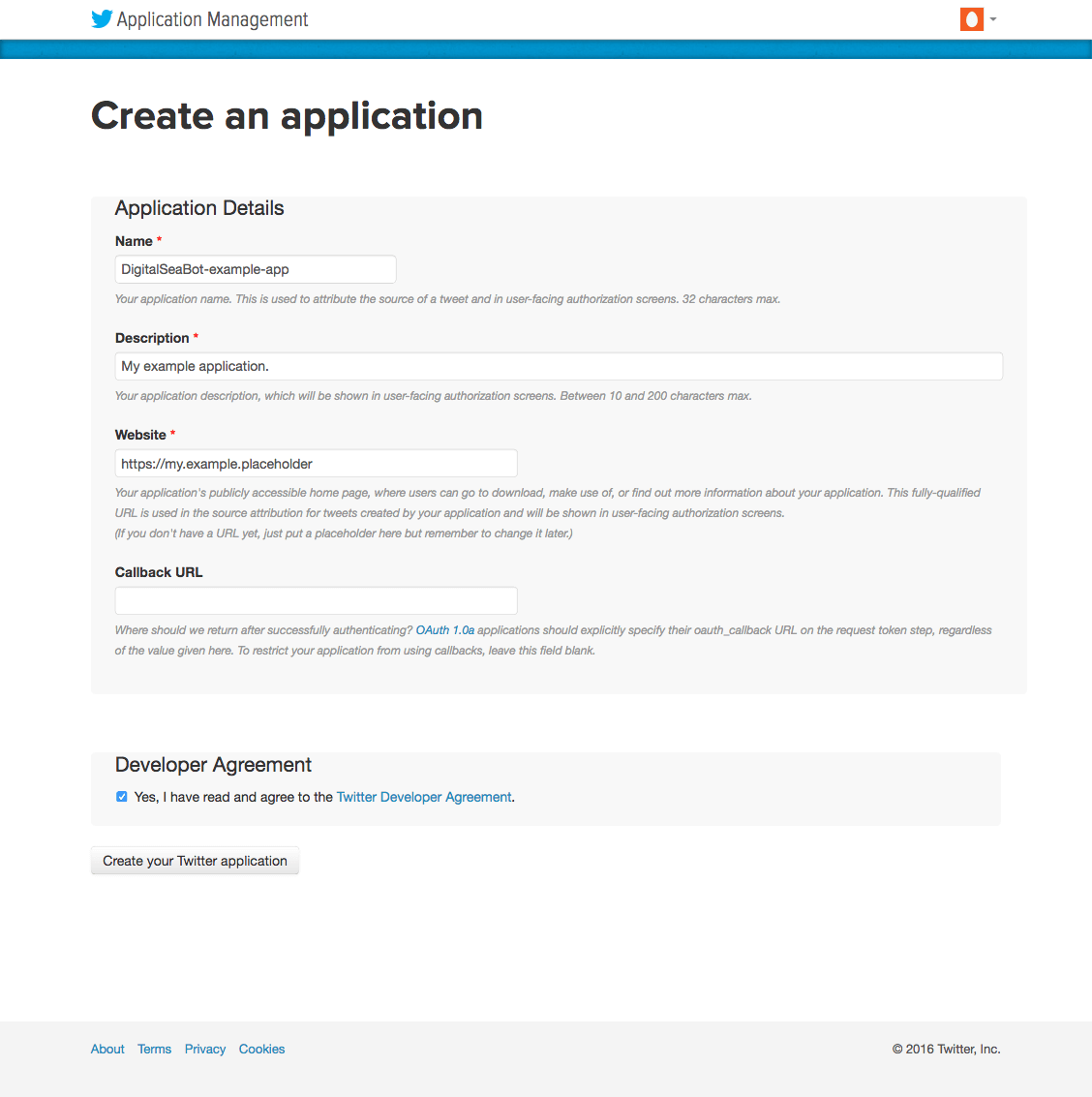
****

Fig 4.3 Creation page of twitter application **Description:**

On this page, you’ll fill out the required fields.

* Name: DigitalSeaBot-example-app
* Description: My example application.
* Website: <https://my.example.placeholder>

Read the [Twitter Developer Agreement](https://dev.twitter.com/overview/terms/agreement-and-policy). If you agree to continue at this point, click the checkbox next to the line that reads, **Yes, I have read and agree to the Twitter Developer Agreement.**

Once you click the **Create your Twitter application** button at the bottom of the page, you’ll receive a confirmation page.

## Step 2 : Modify Your Application’s Permission Level and Generate Your Access Tokens

From the **Details page**, let’s navigate over to the **Permissions** page to ensure that we have the appropriate access level to generate our application keys.

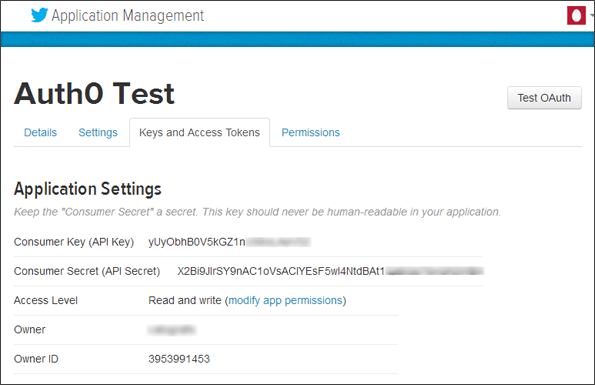


Fig 4.4: Generating keys and Access Tokens

By default, your Twitter app should have Read and Write access. If this is not the case, modify your app to ensure that you have Read and Write access. This will allow your application to post on your behalf. Click on the 'keys and access tokens' tab:

* Copy CONSUMER\_KEY, CONSUMER\_SECRET, ACCESS\_TOKEN, ACCESS\_TOKEN\_SECRET to a file on your computer; never share this file, don't upload it anywhere, don't put it in your open notebook.

**Step 3: install TWARC**

we just need Python on your machine so that your computer understands how to interpret Twarc commands.

1. Download and install [Python](https://www.python.org/downloads/) version 2.7.9.
2. Make sure you have [Pip](https://pip.pypa.io/en/latest/installing.html) installed.
3. Next thing to do is to get the Twarc. This can be done by typing pip install twarc .

## Using Twarc

Twarc will need that information you wrote down in the step above CONSUMER\_KEY, CONSUMER\_SECRET,ACCESS\_TOKEN,ACCESS\_TOKEN\_SECRET

import twarc

consumer\_key = 'your\_consumer\_key'

consumer\_secret = 'your\_consumer\_secret'

access\_token = 'your\_access\_token'

access\_token\_secret = 'your\_access\_token\_secret'

**Chapter 5**

**DESIGN**

**5.1 SYSTEM ARCHITECTURE**

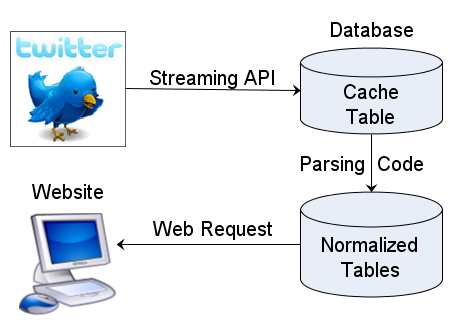
****

Fig 5.1: Architecture of twitter hash tag analysis

**Description:**

From the fig 5.1, first a twitter API should be created using this twitter API data is collected. The collected data is sorted which is known as normalized tables. Then the data is sent to the web page.

**5.2 DATA FLOW DIAGRAMS**

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

Twitter API

Generate keys and access tokens

Use twarc

Access keys&consumer keys

true

Import twarc and write code

Compile code

Enter a hashtag

Displays the hash tag details in sorted order

Fig 5.2: Data flow diagram of hash tag analysis

**5.3 UML DIAGRAMS**

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects**.**

**Goals**

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.

**5.3.1 Use Case Diagram**

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

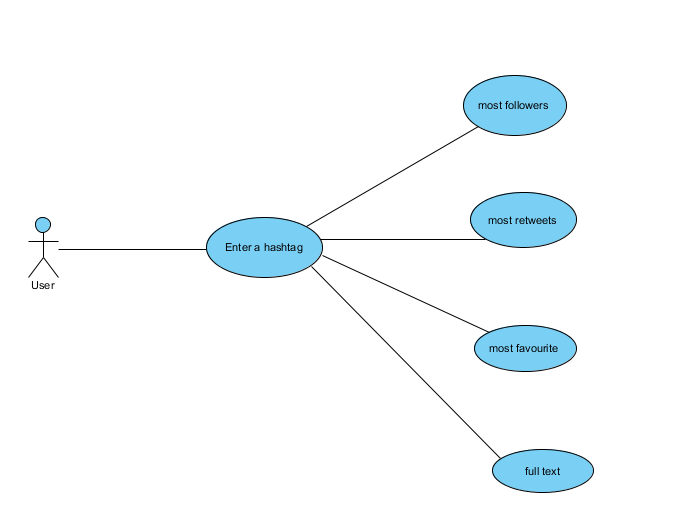
****

Fig 5.3:Usecase diagram of twitter hashtag analysis

**Description**:

The above fig 5.3 represents user, hash tag, most retweets, most favourite, full text usecases.

**5.3.2 Class Diagram**

A class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

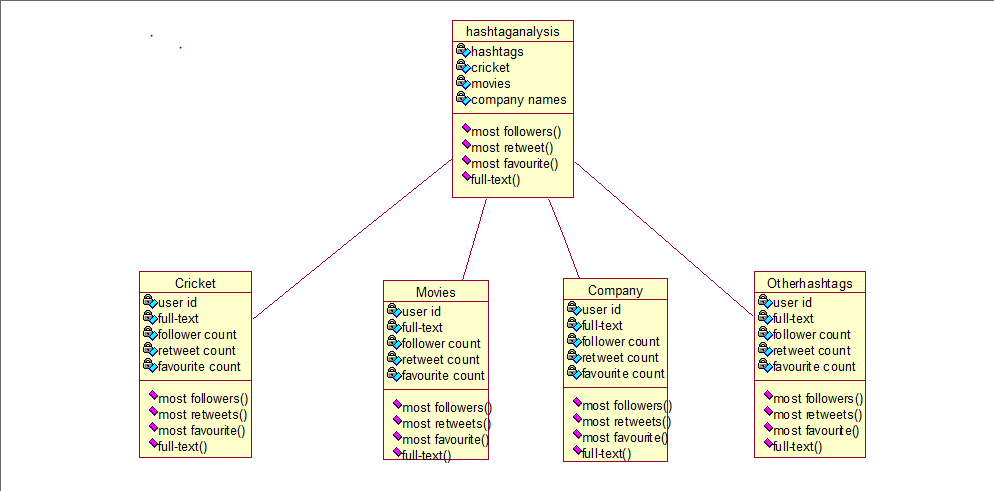
****

Fig 5.4:Class diagram of twitter: hash tag analysis

**Description:**

The above fig 5.4 shows the class diagrams of hashtag analysis, cricket, movies, company, other hashtag attributes and operations and their relation.

**5.3.3 Sequence Diagram**

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

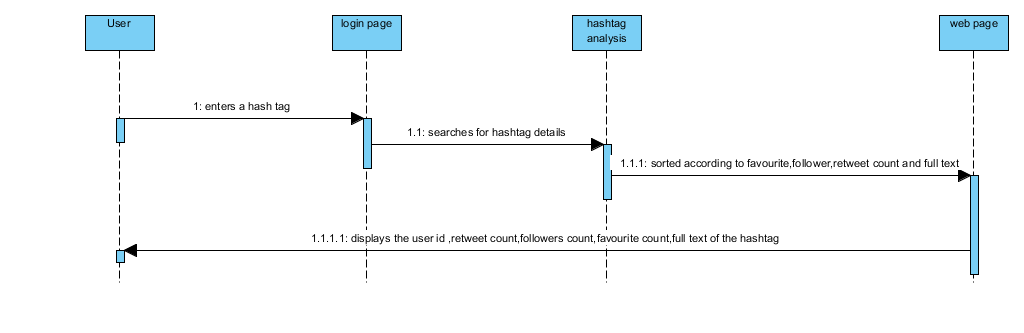


Fig 5.5: Sequence diagram of twitter: hash tag analysis

**Description:**

The above fig 5.5 shows the sequence of actions that are performed in the twitter hashtag analysis.

**Chapter 6**

**IMPLEMENTATION**

Social media marketing is all about engaging with followers and potential customers. The more comments, shares, likes and retweets you are able to get on your social media posts, the greater the visibility will be for your brand.Here’s where hashtags play a crucial role since they drive engagement. Tweets containing hashtags get twice the amount of engagement compared to those without them with 55% more retweets. Adding hashtags to your social media posts makes them more discoverable in search engines and social media platforms.

Twitter hashtag analysis analyzes the information about the hashtag we have entered. It displays userid, retweet count, followers count, favourite count, fulltext of the hash tag. The information is sorted based on 1)most followers, 2)most retweets, 3) most favourite and 4)fulltext.

**MODULES:**

1. Most followers.
2. Most retweets.
3. Most favourite.
4. Full text.

**6.1** **MODULES DESCRIPTION:**

**1.Most followers:**

In this the data is sorted based on the followers count. The User who has the more followers count along with who has used the particular hashtag that we want to search will be displayed first followed by the user with less follower count.

**2.Most favourite:**

In this the data is sorted based on the favourite count. The User who has the most favourite count along with who has used the particular hashtag that we want to search will be displayed first followed by the user with less favourite count.

**3.Most Retweets:**

In this the data is sorted based on the retweet count. The User who has the most retweet count along with who has used the particular hashtag that we want to search will be displayed first followed by the user with less retweet count.

**4.Full Text:**

In this the data is sorted based on the full-text. The User who has entered the more text along with who has used the particular hashtag that we want to search will be displayed first followed by the user with less full text.

**6.2 SAMPLE CODE:**

From import sys,tweepy,csv,re

from textblob import TextBlob

import matplotlib.pyplot as plt

from bottle import run,route,get,post,request,template,static\_file

@route("/sa")

class SentimentAnalysis:

def \_\_init\_\_(self):

self.tweets = []

self.tweetText = []

def DownloadData(self):

# authenticating

consumerKey = 'OGh7cQ6tLezofTRc1r2Gk38Mo'

consumerSecret = 'Jl7XxuagDALyOYQqbbl0M32uDE8f3gAKau8RHp\*\*\*'

accessToken = '1136145916929253376-diwYTpirtPvj6bMc6mmVQJRnd9Rg8W'

accessTokenSecret = 'ASvr17ddDe3jpK6M5TKDcpEtYBGomIq1YetIcEdmtwi7I'

auth = tweepy.OAuthHandler(consumerKey, consumerSecret)

auth.set\_access\_token(accessToken, accessTokenSecret)

api = tweepy.API(auth)

# input for term to be searched and how many tweets to search

searchTerm = input("Enter Keyword/Tag to search about: ")

NoOfTerms = int(input("Enter how many tweets to search: "))

# searching for tweets

self.tweets = tweepy.Cursor(api.search, q=searchTerm, lang = "en").items(NoOfTerms)

# Open/create a file to append data to

#csvFile = open('result.csv', 'a')

# Use csv writer

#csvWriter = csv.writer(csvFile)

# creating some variables to store info

polarity = 0

positive = 0

wpositive = 0

spositive = 0

negative = 0

wnegative = 0

snegative = 0

neutral = 0

# iterating through tweets fetched

for tweet in self.tweets:

#Append to temp so that we can store in csv later. I use encode UTF-8

self.tweetText.append(self.cleanTweet(tweet.text).encode('utf-8'))

# print (tweet.text.translate(non\_bmp\_map)) #print tweet's text

analysis = TextBlob(tweet.text)

# print(analysis.sentiment) # print tweet's polarity

polarity += analysis.sentiment.polarity # adding up polarities to find the average later

if (analysis.sentiment.polarity == 0): # adding reaction of how people are reacting to find average later

neutral += 1

elif (analysis.sentiment.polarity > 0 and analysis.sentiment.polarity <= 0.3):

wpositive += 1

elif (analysis.sentiment.polarity > 0.3 and analysis.sentiment.polarity <= 0.6):

positive += 1

elif (analysis.sentiment.polarity > 0.6 and analysis.sentiment.polarity <= 1):

spositive += 1

elif (analysis.sentiment.polarity > -0.3 and analysis.sentiment.polarity <= 0):

wnegative += 1

elif (analysis.sentiment.polarity > -0.6 and analysis.sentiment.polarity <= -0.3):

negative += 1

elif (analysis.sentiment.polarity > -1 and analysis.sentiment.polarity <= -0.6):

snegative += 1

# Write to csv and close csv file

#csvWriter.writerow(self.tweetText)

#csvFile.close()

# finding average of how people are reacting

positive = self.percentage(positive, NoOfTerms)

wpositive = self.percentage(wpositive, NoOfTerms)

spositive = self.percentage(spositive, NoOfTerms)

negative = self.percentage(negative, NoOfTerms)

wnegative = self.percentage(wnegative, NoOfTerms)

snegative = self.percentage(snegative, NoOfTerms)

neutral = self.percentage(neutral, NoOfTerms)

# finding average reaction

polarity = polarity / NoOfTerms

# printing out data

print("How people are reacting on " + searchTerm + " by analyzing " + str(NoOfTerms) + " tweets.")

print()

print("General Report: ")

if (polarity == 0):

print("Neutral")

elif (polarity > 0 and polarity <= 0.3):

print("Weakly Positive")

elif (polarity > 0.3 and polarity <= 0.6):

print("Positive")

elif (polarity > 0.6 and polarity <= 1):

print("Strongly Positive")

elif (polarity > -0.3 and polarity <= 0):

print("Weakly Negative")

elif (polarity > -0.6 and polarity <= -0.3):

print("Negative")

elif (polarity > -1 and polarity <= -0.6):

print("Strongly Negative")

print()

print("Detailed Report: ")

print(str(positive) + "% people thought it was positive")

print(str(wpositive) + "% people thought it was weakly positive")

print(str(spositive) + "% people thought it was strongly positive")

print(str(negative) + "% people thought it was negative")

print(str(wnegative) + "% people thought it was weakly negative")

print(str(snegative) + "% people thought it was strongly negative")

print(str(neutral) + "% people thought it was neutral")

self.plotPieChart(positive, wpositive, spositive, negative, wnegative, snegative, neutral, searchTerm, NoOfTerms)

def cleanTweet(self, tweet):

# Remove Links, Special Characters etc from tweet

return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t]) | (\w +:\ / \ / \S +)", " ", tweet).split())

# function to calculate percentage

def percentage(self, part, whole):

temp = 100 \* float(part) / float(whole)

return format(temp, '.2f')

def plotPieChart(self, positive, wpositive, spositive, negative, wnegative, snegative, neutral, searchTerm, noOfSearchTerms):

labels = ['Positive [' + str(positive) + '%]', 'Weakly Positive [' + str(wpositive) + '%]','Strongly Positive [' + str(spositive) + '%]', 'Neutral [' + str(neutral) + '%]',

'Negative [' + str(negative) + '%]', 'Weakly Negative [' + str(wnegative) + '%]', 'Strongly Negative [' + str(snegative) + '%]']

sizes = [positive, wpositive, spositive, neutral, negative, wnegative, snegative]

colors = ['yellowgreen','lightgreen','darkgreen', 'gold', 'red','lightsalmon','darkred']

patches, texts = plt.pie(sizes, colors=colors, startangle=90)

plt.legend(patches, labels, loc="best")

plt.title('How people are reacting on ' + searchTerm + ' by analyzing ' + str(noOfSearchTerms) + ' Tweets.')

plt.axis('equal')

plt.tight\_layout()

plt.show()

if \_\_name\_\_== "\_\_main\_\_":

sa = SentimentAnalysis()

sa.DownloadData()

run(debug=True, reloader=True)

**hashtag.py**

from bottle import run,route,get,post,request,template,static\_file

from twarc import Twarc

import pandas as pd

t=Twarc("6EtpP94ELaRdXO3ItbUcOb5e3","pYghu8gl5FRJF16ZKR8GtsgPiJOCjwq5Hc76inqzDmrUZRpyCj","1002781398531391488-nKedW6KnV0XZvOpKegwAsb2DCLFZKT","3YzUdH74zhKf3T4ZvJF5YJsde34Aa303IiQ7hl05ko2Gf")

twdata = None

@route("/<hashtag>")

def search(has):

global twdata,data2,data4,data6

c=0

news=[]

for tweet in t.search(has):

tweet['followers\_count'] = tweet['user']['followers\_count']

news.append(tweet)

if c>100:

break

c=c+1

data = pd.DataFrame(news)

data['favorite\_count'] = data['favorite\_count'].apply(pd.to\_numeric)

data = data.sort\_values(by='favorite\_count', ascending=False)

twdata = data

data1 = data.sort\_values(by='followers\_count', ascending=False)

data2 = data1

data3 = data.sort\_values(by='retweet\_count', ascending=False)

data4 = data3

data5 = data.sort\_values(by='full\_text', ascending=False)

data6 = data5

return template('temp', data=twdata)

@route('/favs')

def favs():

global twdata

count = twdata.sort\_values(by='favorite\_count',ascending=False)

return template('favs',data=twdata)

@route('/user')

def user():

global data2

user = twdata.sort\_values(by='followers\_count',ascending=False)

return template('favs',data=data2)

@route('/retweets')

def retweets():

global data4

user = twdata.sort\_values(by='retweet\_count',ascending=False)

return template('favs',data=data4)

@route('/fulltext')

def fulltext():

global data6

user = twdata.sort\_values(by='full\_text',ascending=False)

return template('favs',data=data6)

@get("/proj")

def hello():

return template("proj", data=twdata)

@post("/get\_details")

def do\_get():

has=request.forms.get("hashtag")

return search(has)

run(host="localhost",reloader="True",port="8002")

**proj.html**

<!doctype html>

<html>

<head>

<title>twitter</title>

<style>

body{

background-image:url("https://i.pinimg.com/originals/fc/aa/23/fcaa23ba0b59b0abd2359f2c3f8fcc5a.jpg");

background-size: 1300px 800px;

background-repeat:no-repeat;

}

#rcorners3{

border-radius: 80px 0px;

background-image:url("http://www.hdwallpapers10.com/wp-content/uploads/2017/05/Black%20and%20White%20abstract%20Background%20Full%20HD-623x623.png");

padding: 20px;

width: 500px;

height: 200px;

opacity:0.8;

}

</style>

</head>

<body><b>

<h1 align="center" style="color:royalblue">Hashtag Analysis</h1></b>

<form id=rcorners3 method="post" action="/get\_details">

<img src="http://realmomentsphotography.com.au/wp-content/uploads/2016/07/twitter.png" height="100px" width="100px">

<p align="center" style="color:deeppink"><b>Hashtag:</b>

<input type="text" placeholder="enter hashtag" name="hashtag" ><br>

<br>

<input type="submit" value="Search"></p>

</form>

</body>

</html>

**temp.html**

<!DOCTYPE html>

<html>

<head>

<style>

body{

background:url("https://proxy.duckduckgo.com/iu/?u=http%3A%2F%2F1.bp.blogspot.com%2F-yPpg6IOFmQI%2FVimWK4Vus\_I%2FAAAAAAAAGAY%2Fxg6RvIoiN-g%2Fs1600-r%2F541.png&f=1");

background-size:1300px 660px;

background-repeat:no-repeat;

}

ul{margin:0px;

padding:0px;

}

ul li a{

text-decoration:none;

color:red;

display:block;

}

ul li

{

float:left;

width:250px;

height:50px;

background-color:white;

color:white;

font-size:15px;

line-height:40px;

text-align:center;

}

ul li a:hover{

background-color:orange;

}

ul li ul li{

display:none;

}

ul li:hover ul li

{

display:block;

}

</style>

</head>

<body>

<div>

<ul>

<li><a href="/favs">Most favorite</a></li>

<li><a href="/user">Most followers</a></li>

<li><a href="/retweets">Most Retweet</a></li>

<li><a href="/fulltext">Full Text</a></li>

</ul>

</div><br><br><br><br><br><br><br><br><br><br><br><br><br>

</body>

</html>

**favs.html**

<!DOCTYPE html>

<html>

<head>

<style>

body{

background:url("https://proxy.duckduckgo.com/iu/?u=http%3A%2F%2Fwww.ie-wallpapers.com%2Fdata%2Fout%2F186%2F35942141-plain-background-images.jpg&f=1");

background-size:1300px 660px;

background-repeat:repeat;

}

</style>

</head>

<body>

</body>

</html>

%for index, t in data.iterrows():

<h3> Full Text: </h3>{{ t['full\_text']}}<br>

<h3> Retweet count:</h3>{{ t['retweet\_count']}}<br>

<h3> Favorite count:</h3>{{ t['favorite\_count']}}<br>

<h3> User id:</h3> {{ t['id']}}<br>

<h3> Follower count:</h3>{{ t['user']['followers\_count']}}<br>

<br>

<h3>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------</h3>

<br>

%end

**Chapter 7**

**TESTING**

**7.1 INTRODUCTION TO TESTING**

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that theSoftware system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

**7.2 TYPES OF TESTS**

**Unit Testing**

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

**Integration Testing**

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

**Functional Test**

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

**System Test**

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

**White Box Testing**

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

**Black Box Testing**

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box. It cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

**7.3 TESTING TECHNIQUES**

**Unit Testing**

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

**Test strategy and approach**

Field testing will be performed manually and functional tests will be written in detail.

**Test objectives**

* All field entries must work properly.
* Pages must be activated from the identified link.
* The entry screen, messages and responses must not be delayed.

**Features to be tested**

* Verify that the entries are of the correct format
* No duplicate entries should be allowed
* All links should take the user to the correct page.

**Integration Testing**

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level -interact without error.

**Test Results:**

All the test cases mentioned above passed successfully. No defects encountered.

**Acceptance Testing**

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user.

**7.4 Test cases:**

**Hashtag page**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test Case**  **No** | **Test Data** | **Test**  **Description** | **Expected**  **Result** | **Actual**  **Result** | **Status** |
| TC – 01 | Hashtag | Valid hashtag | Hash tag results | Valid details | pass |
| TC-01 | Hash tag | Invalid hashtag | errors | Invalid details | fail |

**Table 7.4.1 Hashtag page**

**Hashtag Analysis page**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test Case**  **No** | **Test Data** | **Test**  **Description** | **Expected**  **Result** | **Actual**  **Result** | **Status** |
| TC – 02 | Hashtag analysis | All the details of the hash tag displayed. | Hash tag results | Valid details | pass |
| TC-02 | Hashtag analysis | All the details of hash tag are not displayed | errors | Invalid details | fail |

**Table 7.4.2 Hashtag Analysis page**

**Test Results:**

All the test cases mentioned above passed successfully. No defects encountered.

**7.5 SCREENSHOTS**

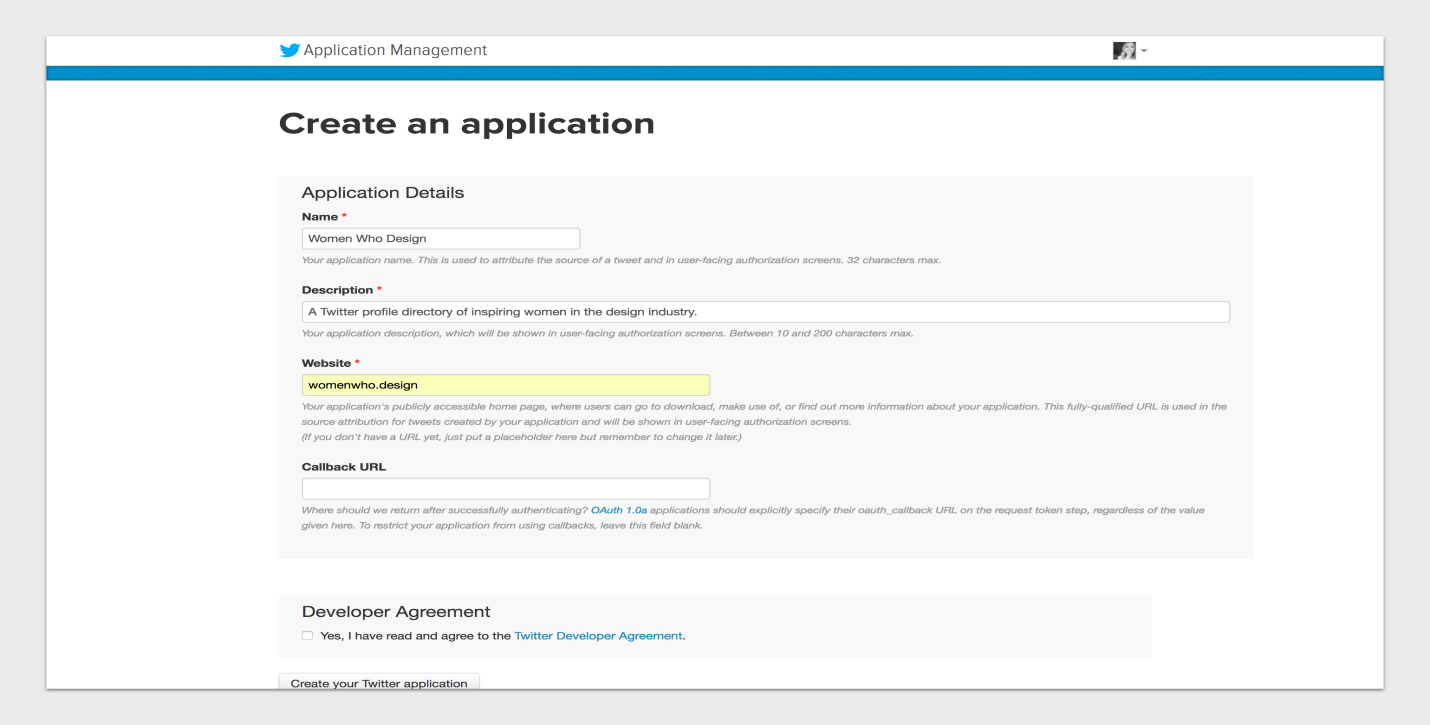


Fig no 7.1:Twitter API

**Description for twitter API:**

Twitter allows you to interact with its data i.e tweets and s everal attributes about tweets using Twitter API

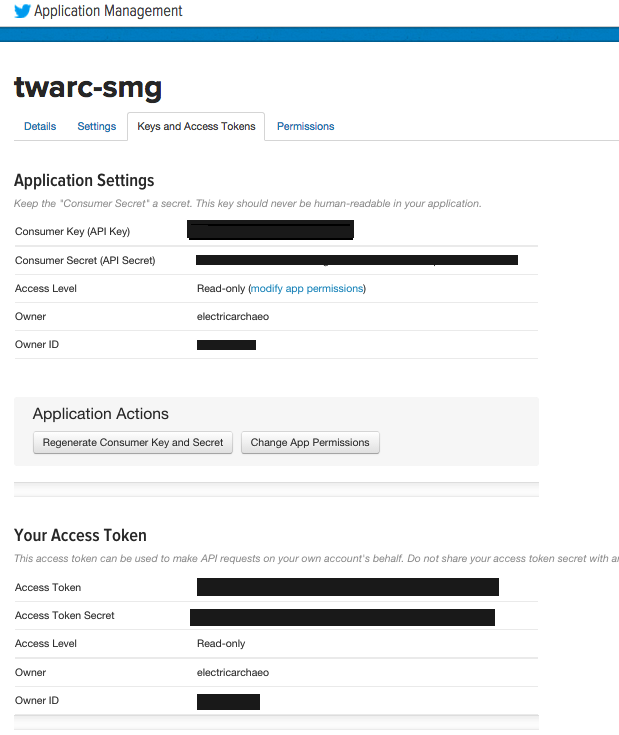
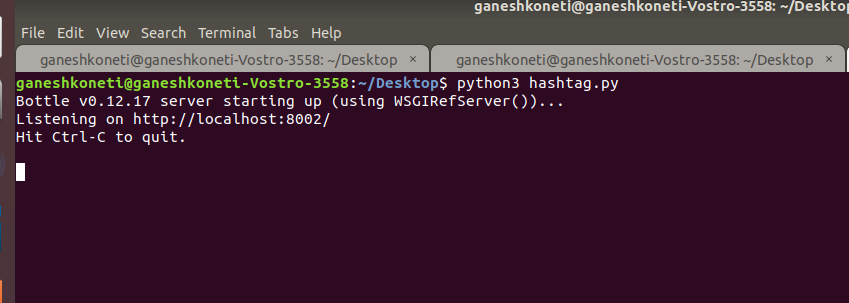


Fig no 7.2:TWARC

**Description for twarc:**

twarcis a command line tool and Python library for archiving Twitter JSON data.

 Fig no 7.3 Compiling the code

**Description for compiling the code:**

Here the code is compiled .The code is compiled as python hashtag.py



Fig no 7.4:hashtag analysis

**Description for the twitter hashtag analysis :**

A web page is displayed where we have to enter the hashtag to be searched.

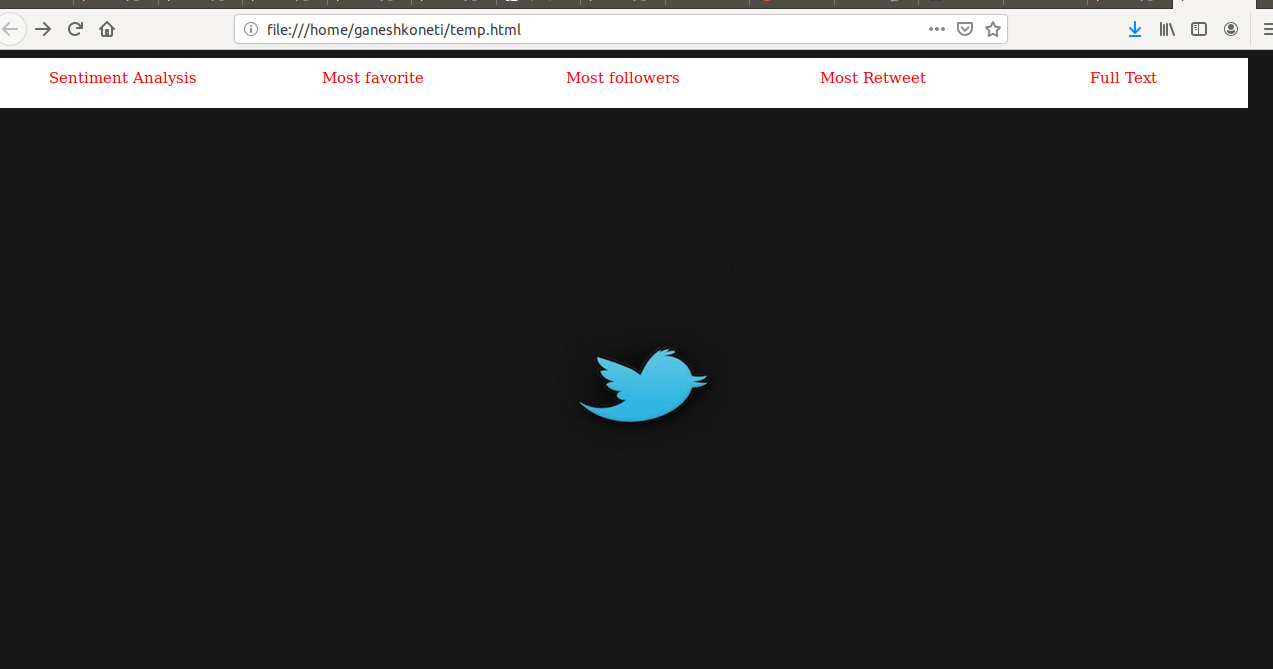


Fig no 7.5:Data Access Request page

**Description for Data Access Request Page:**

The data is analysed for the particular hashtag and is sorted according to the full-text, retweet count, favourite count, follower count

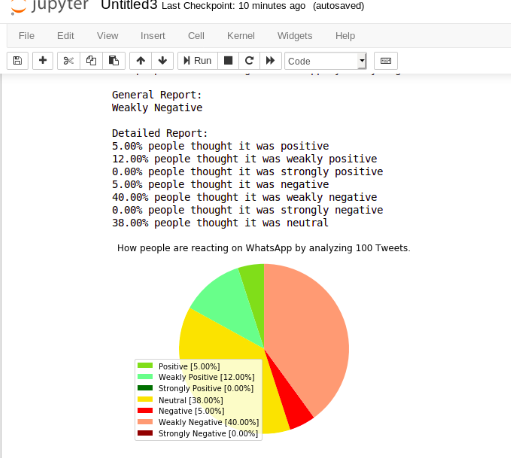
****

Fig no 7.6:pie chart for the hashtag

**Description for the twitter hashtag :**

The data is analysed for the particular hashtag and is sorted according to the positive, weakly positive, strongly positive, neutral, negative, weakly negative, strongly negative.

**Chapter 8**

**CONCLUSION**

Hashtags are now widespread on social networks, and the term is recognized in the dictionary. Using hashtags puts pertinent information at your fingertips, allowing you to spend less time searching for information..Spend just a few minutes researching popular hashtags and you’re sure to find what works for you.

Hashtags simplify the process-searching a hash tag pulls results for each post using that hash tag. Using a hash tag helps you reach your target audience, and likewise makes it easier for others to find your information. Hash tag compel an action-when a user sees a post that is of interest, they will likely spend time looking through content brought up by the hashtag.

Hashtags evolve-Hashtags are being used by more and more platforms, impacting the amount of information put directly in front of social media users. Hashtag reward the distinctive-Hashtags make finding information easier for social media users. A unique hash tag makes your message stand out to the users who find the hash tag valuable.

In this project we have done the hash tag analysis which will be very helpful for the business people to know the followers count, favourite count, retweet count, full text and user id of that particular hash tag.

**Chapter 9**

**Future Scope**

Trend Analysis. Trends as a useful way for people to find out what topics are being talked about around the world, right now. Trends can be viewed globally i.e worldwide or locally i.e. specific countries and metropolitan location .Twitter uses an algorithms that determines which topics are “trending” in the location which is selected. This algorithm identifies topics that are immediately popular, rather than topics that have been popular for a while or on a daily basis, to help people discover the hottest emerging topics of discussion on twitter. some topics include a #(pound sign) before the word or phrase, this is called a hash tag.

**BIBLIOGRAPHY**

[1] Abel, F., Celik, I., Houben, G., & Siehndel, P. (2011). Leveraging the Semantics of Tweets for Adaptive Faceted Search on Twitter. Proceedings of the 10th International Semantic Web Conference (pp. 1–17).

[2] Abel, F., Herder, E., Houben, G., & Henze, N. (2011). Cross-system User Modeling and Personalization on the Social Web. User Modeling and User-Adapted Interaction, 22(3), 1–42.

Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting Spammers on Twitter. Collaboration, electronic messaging, anti-abuse and spam conference (CEAS).

Bruns, A., & Burgess, J. (2011). The Use of Twitter Hashtags in the Formation of Ad Hoc Publics(pp.25–27).Retrievedfrom<http://eprints.qut.edu.au/46515/1/The_Use_of_Twitter_Hashtags_in_the_Formation_of_Ad_Hoc_Publics_(final).pdf>

[3] Cheong, M., & Lee, V. (2010). Dissecting Twitter: A Review on Current Microblogging Research and Lessongs from Related Fields. From Sociology to Computing in Social Networks (pp. 343–362). Springer Vienna.

[4] Cheong, M., & Ray, S. (2011). A Literature Review of Recent Microblogging Developments(pp.1–43).Retrievedfrom http://www.csse.monash.edu.au/publications/2011/tr-2011-263-full.pdf

Dann, S. (2010). Twitter content classification. First Monday, 15(12). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/2745/2681>

[5] Gwet, K. L. (2012). Benchmarking Inter-Rater Reliability Coefficients. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters (3rd Editio., pp. 121–128). Advanced Analytics Press.

[6] Honeycutt, C., & Herring, S. C. (2009). Beyond Microblogging : Conversation and Collaboration via Twitter. System Sciences, 2009. HICSS’09. 42nd Hawaii International Conference on (pp. 1–10).

[7] Horn, C. (2010). Analysis and Classification of Twitter messages. Retrieved from <http://know-center.tugraz.at/wp-content/uploads/2010/12/Master-Thesis-Christopher-Horn.pdf>

[8] Huang, J., Thornton, K. M., & Efthimiadis, E. N. (2010). Conversational tagging in twitter. Proceedings of the 21st ACM conference on Hypertext and hypermedia - HT ’10 (p. 173). New York, New York, USA: ACM Press. doi:10.1145/1810617.1810647

[9] Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter Power : Tweets as Electronic Word of Mouth. Journal of the American Society for Information Science and Technology, 60(11), 2169–2188. doi:10.1002/asi

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why We Twitter : Understanding Microblogging. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (pp. 56–65).

[10] Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features.

[11] Springer. Krishnamurthy, B., Gill, P., & Arlitt, M. (2008). A Few Chirps About Twitter. Proceedings of the first workshop on Online social networks (pp. 19–24).

[12] Lin, Y., Margolin, D., Keegan, B., Baronchelli, A., & Lazer, D. (2012). # Bigbirds Never Die : Understanding Social Dynamics of Emergent Hashtags.

[13] Liu, W. (2013). How Twitter Connects to Information Sources : Proceedings of the Annual Conference of International Communication Association.

[14] Magnani, M., Montesi, D., Nunziante, G., & Rossi, L. (2011). Conversation Retrieval from Twitter. Proceedings of the 33rd European Conference on IR Research, ECIR 2011 (pp. 780–783).

[15] Naaman, M., Boase, J., & Lai, C. (2010). Is it Really About Me ? Message Content in Social Awareness Streams. Proceedings of the 2010 ACM conference on Computer supported cooperative work (pp. 189–192).

[16] Neuendorf, K. A. (2002). The content analysis guidebook. Sage.

1. [↑](#footnote-ref-2)